MQB7046: MODELLING PUBLIC HEALTH DATA

Semester 2, Session 2023/2024

CONTINUOUS ASSESSMENT 2

Please read the instructions carefully. Failure to comply with any of these instructions may result in penalty of marks or grading of this assessment.

1. This assignment represents 30% of the evaluation for this course.
2. Answer ALL questions in English.
3. This assignment is "open book," which means you are permitted to use any materials handed out in class, your own notes from the course and textbooks.
4. All analyses should be conducted using the Python software. All codes and outputs need to be interpreted.
5. Submission instructions:
   - For students found plagiarising or commit any forms of academic misconducts, they will be penalised and given a zero (fail).
   - Upload in a folder the following information into SPECTRUM; 1) your jupyter notebook (with output), 2) any additional report (optional)
   - All documents must be named according to the Matrix_number
   - Submit your files via SPeCTRUM before the due date on **15/5/2022 11.59pm.** Failure to submit your assignment on the stated due date and time will be given a zero.

QUESTION:

Data were collected for 1,200 men and women in a study on mental health. In this cross-sectional study, all participants were asked a series of 20 questions about depressive symptoms, and a sum score of depressive symptoms was constructed. All other variables were also collected during interviews in the participants' homes. The dataset consists of the following information:

Variable	Coding
age	age in years
sex	sex: 1=male 2=female
educ	education:1=primary,2=vocational,3=secondary, 4=university
married	marital status: 0=married 1=unmarried
deprivation	material deprivation: 1=yes 0=no

bmi	body mass index in kg/m2
sbp	systolic blood pressure in mmHg
alcohol	problem drinking: 1=yes 0=no
smoke	current smoking: 1=yes 0=no
depscore	depression score: score ranges from 0-60

Score of depressive symptoms (CES-D score) was constructed from 20 questions. The answers for each question are on a scale of 0-3, and the overall score ranges from 0-60. Score 0 means no depressive symptoms, score 60 means high score of depressive symptoms. Based on evidence from previous studies in various populations, respondents can be categorised into the following groups: a) not depressed (0–9 points), b) mildly depressed (10–15 points), c) moderately depressed (16–24 points), or d) severely depressed (more than 25 points). In other studies, a score of less than 16 indicates no depression, 16 and more indicates a clinically relevant depression.

The investigators are interested in the relationship between material deprivation of individuals and depression. It has been hypothesized that people who are more deprived are more likely to develop depression/more depressive symptoms than people in better material circumstances.

Instructions:

Perform any relevant analyses and write a report that describe your analysis steps and summarise your findings.

a) State the related research question, objective and hypothesis based on the information provided above.
b) Summarise all variables available in the dataset.
c) Based on the information provided in (a), perform any required analyses to facilitate drawing relevant conclusions. Undertake any necessary data cleaning and preprocessing as needed. Use appropriate statistical methods or techniques and specify any assumptions necessary for the analysis. If warranted, provide justification for the chosen approach.
   - Decide how you would like to analyse your dependent variable (continuous or categorical).
   - Evaluate the association between possible risk factors and depression using non-regression and regression analysis.
   - Also decide what variables that you need to consider as possible confounders or effect modifiers? If yes/no, provide your justifications
   - Perform any additional analysis required to test these assumptions.
d) Summarise your results in appropriate tabular format and comment on your findings. State your conclusion.

(30 marks)